

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-259648

(P2000-259648A)

(43) 公開日 平成12年9月22日 (2000.9.22)

(51) Int.Cl.

識別記号

F I

テーマコード(参考)

G 0 6 F 17/30

G 0 6 F 15/40

3 7 0 A 5 B 0 7 5

15/419

3 2 0

審査請求 未請求 請求項の数16 O L (全 16 頁)

(21) 出願番号

特願平11-58791

(22) 出願日

平成11年3月5日 (1999.3.5)

(71) 出願人 000005049

シャープ株式会社

大阪府大阪市阿倍野区長池町22番22号

(72) 発明者 ▲吉▼田 広市

大阪府大阪市阿倍野区長池町22番22号 シ

ャープ株式会社内

(74) 代理人 100064746

弁理士 深見 久郎

Fターム(参考) 5B075 KK13 KK34 MM11 ND03 ND16

ND20 ND35 NK02 NK54 NS10

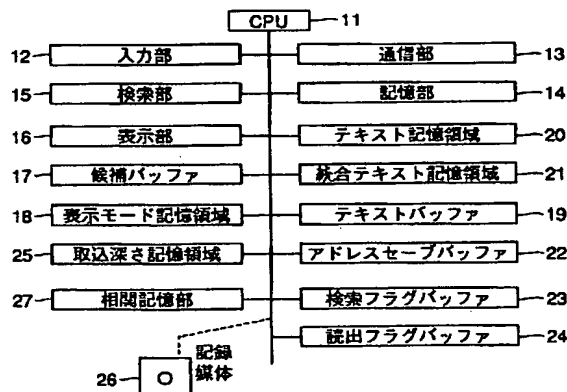
PR04 UU06

(54) 【発明の名称】 文章検索装置および方法ならびに文章検索プログラムを記録したコンピュータで読取り可能な記録媒体

(57) 【要約】

【課題】 検索効率を高めることのできる文章検索装置を提供する。

【解決手段】 CPU 11の制御の下、通信部13を介して文章間にリンクが張られたハイパーテキストを構成する文章を含む複数のデータを収集する。そして、収集されたデータについてハイパーテキストのリンク関係から判断してリンクされている複数の文章が1つの文章に統合されて統合文章記憶領域21に記憶される。その後、入力部12から1つ以上の検索文字列が指定して入力されると、記憶領域21の統合文章から検索文字列が検索部15により検索されて、その検索結果が表示部16に表示される。



BEST AVAILABLE COPY

## 【特許請求の範囲】

【請求項1】 各データ内で相互に関連付けるためのリンク情報に従い関連付けられた文章データを含む複数のデータを記憶する記憶手段と、

前記記憶手段中の前記複数のデータを前記リンク情報に従い統合処理して1つ以上の統合文章データに生成するための統合手段と、

1つ以上の検索文字列データを指定する指定手段と、前記統合手段により統合された前記1つ以上の統合文章データのそれぞれにおいて、前記指定手段により指定された前記1つ以上の検索文字列データが含まれるか否かを検索する検索手段と、

前記検索手段の検索結果を表示する表示手段とを備えた、文章検索装置。

【請求項2】 前記リンク情報を記憶するリンク情報記憶手段をさらに有して、

前記統合手段は、

前記リンク情報記憶手段中の前記リンク情報に基づいて前記複数のデータにおいて前記関連付けがループしているか検出するループ検出手段をさらに有して、前記ループ検出手段により前記複数のデータにおいてループしていると検出された前記データは前記統合処理の対象としないことを特徴とする、請求項1に記載の文章検索装置。

【請求項3】 前記統合手段は、前記ループ検出手段によりループしていると検出された場合には、ループしている前記関連付けにおける1つ以上の部分で前記関連付けが無しとされた後の前記リンク情報に基づいて、前記複数のデータについて前記統合処理をすることを特徴とする、請求項2に記載の文章検索装置。

【請求項4】 前記統合手段は、前記リンク情報記憶手段中の前記リンク情報に基づいて前記複数のデータにおいて一方の前記文章データから他方の前記文章データへの前記関連付けを判定する関連付け判定手段をさらに有し、

前記関連付け判定手段により前記一方文章データ中の複数の単語のそれぞれから前記他方文章データ中の複数の単語それぞれに関連付けされていると判定された場合には、前記統合手段は前記他方文章データを前記関連付けに基づいて分割して、分割後の文章データのそれぞれを用いて前記統合処理をすることを特徴とする、請求項2または3に記載の文章検索装置。

【請求項5】 前記関連付け判定手段により、前記一方文章データ中の複数の単語から前記他方文章データ中の同じ単語に関連付け去れていると判定された場合には、前記統合手段は前記他方文章データを分割せずに前記統合処理をすることを特徴とする、請求項4に記載の文章検索装置。

【請求項6】 前記統合処理において、前記文章データが統合される場合には、前記関連付けにおける関連元の前記文章データ中に対応の関連先の前記文章データが段

落を下げた取込まれることを特徴とする、請求項1ないし5のいずれかに記載の文章検索装置。

【請求項7】 前記統合処理において、前記文章データが統合される場合に、前記関連付けにおける関連元の前記文章データ中に対応する関連先文章データの取込位置の文字列と該関連先文章データの先頭文字列が同じであるときは、一方の文字列が省略されることを特徴とする、請求項1ないし6のいずれかに記載の文章検索装置。

10 【請求項8】 前記表示手段において、前記検索結果である前記統合文章データはアウトライン機能モードで表示されることを特徴とする、請求項1ないし7のいずれかに記載の文章検索装置。

【請求項9】 前記統合処理において、関連先の前記データが前記文章データでない場合には、該データは処理対象とされないことを特徴とする請求項1ないし8のいずれかに記載の文章検索装置。

20 【請求項10】 前記統合処理において、関連先の前記データが前記文章データでない場合には、該データを識別するために予め付与された文字列データを用いて処理されることを特徴とする、請求項1ないし8のいずれかに記載の文章検索装置。

【請求項11】 前記記憶手段中の前記複数のデータは、ハイパーテキストを構成するデータであることを特徴とする、請求項1ないし10のいずれかに記載の文章検索装置。

【請求項12】 各データ内で相互に関連付けるためのリンク情報に従い関連付けられた文章データを含む複数のデータを記憶する記憶手段と、

30 1つ以上の検索文字列データを指定する指定手段と、前記記憶手段中の前記複数のデータ中から前記指定手段により指定された前記検索文字列データを検索する検索手段と、前記検索手段による検索結果を表示する表示手段とを備え、

前記検索手段は、前記データの検索中に前記リンク情報があれば、該データの検索を中断し該リンク情報に従う関連先の前記データを検索して検索終了すれば、該リンク情報に従う関連元の前記データの検索を再開しながら、前記1つ以上の検索文字列データが検索された前記文章データを逐次統合して統合文章データに生成することを特徴とする、文章検索装置。

【請求項13】 各データ内で相互に関連付けるためのリンク情報に従い関連付けられた文章データを含む複数のデータを対象にした文章検索方法であって、前記複数のデータを前記リンク情報に従い統合処理して1つ以上の統合文章データに生成するための統合ステップと、

50 1つ以上の検索文字列データを指定する指定ステップと、

前記統合ステップにより統合された前記1つ以上の統合文章データのそれぞれにおいて、前記指定ステップにより指定された前記1つ以上の検索文字列データが含まれるか否か検索する検索ステップと、  
前記検索ステップの検索結果を表示する表示ステップとを備えた、文章検索方法。

【請求項14】 各データ内で相互に関連付けるためのリンク情報に従い関連付けられた文章データを含む複数のデータを対象にした文章検索方法であって、

1つ以上の検索文字列データを指定する指定ステップ 10 と、

前記複数データ中から前記指定ステップにより指定された前記検索文字列データを検索する検索ステップと、  
前記検索ステップによる検索結果を表示する表示ステップとを備え、

前記検索ステップは、前記データの検索中に前記リンク情報があれば、該データの検索を中断し該リンク情報に従う関連先の前記データを検索して検索終了すれば、該リンク情報に従う関連元の前記データの検索を再開しながら、前記1つ以上の検索文字列データが検索された前記文章データを逐次統合して統合文章データに生成することを特徴とする、文章検索方法。 20

【請求項15】 各データ内で相互に関連付けるためのリンク情報に従い関連付けられた文章データを含む複数のデータを対象にした文章検索方法をコンピュータに実行させるための文章検索プログラムを記録したコンピュータで読取り可能な記録媒体であって、

前記文章検索方法は、  
前記複数のデータを前記リンク情報に従い統合処理して 30 1つ以上の統合文章データに生成するための統合ステップと、

1つ以上の検索文字列データを指定する指定ステップと、

前記統合ステップにより統合された前記1つ以上の統合文章データのそれぞれにおいて、前記指定ステップにより指定された前記1つ以上の検索文字列データが含まれるか否か検索する検索ステップと、

前記検索ステップの検索結果を表示する表示ステップとを備えた、文章検索プログラムを記録したコンピュータで読取り可能な記録媒体。

【請求項16】 各データ内で相互に関連付けるためのリンク情報に従い関連付けられた文章データを含む複数のデータを対象にした文章検索方法をコンピュータに実行させるための文章検索プログラムを記録したコンピュータで読取り可能な記録媒体であって、

前記文章検索方法は、  
各データ内で相互に関連付けるためのリンク情報に従い 40 関連付けられた文章データを含む複数のデータを対象にした文章検索方法であって、

1つ以上の検索文字列データを指定する指定ステップ 50

と、

前記複数データ中から前記指定ステップにより指定された前記検索文字列データを検索する検索ステップと、  
前記検索ステップによる検索結果を表示する表示ステップとを備え、

前記検索ステップは、前記データの検索中に前記リンク情報があれば、該データの検索を中断し該リンク情報に従う関連先の前記データを検索して検索終了すれば、該リンク情報に従う関連元の前記データの検索を再開しながら、前記1つ以上の検索文字列データが検索された前記文章データを逐次統合して統合文章データに生成することを特徴とする、文章検索プログラムを記録したコンピュータで読取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明はハイパーテキストなど、特定のキーワードなどで有機的に結び合わせて作成される文章を検索対象とする文章検索装置および方法ならびに文章検索プログラムを記録したコンピュータで読取り可能な記録媒体に関し、特に、リンクされている複数の文章を1つに統合して検索する文章検索装置および方法ならびに文章検索プログラムを記録したコンピュータで読取り可能な記録媒体に関する。

【0002】なお、ここでリンクを張るとは、相互に関連付けることをいう。またリンク情報とはこの関係付けを示す情報を指す。ハイパーテキストとは、音声データ、画像データ、文章データなど同士をデータ中でリンクを張ることによって有機的に結合して作成された文章をいう。

【0003】

【従来の技術および発明が解決しようとする課題】近年、文章の検索システムとして種々のものが実用化されている。その1つに、インターネットの発達で各種ホームページの情報を収集するサーチエンジン（検索システム）が開発されている。このサーチエンジンでは、ホームページが階層化されて文字列が分散していると検索したい文字列が別々の文章に分かれて存在するため文字列が完全に一致せずに、検索できない。逆に、検索する文字列を短く区切ると不要なホームページが大量に検索抽出されてしまう。 40

【0004】全部を検索対象とするいわゆるフルテキストサーチ技術は、インターネットのサーチエンジンでよく使われている。しかし、検索文字列が一致しているかどうかは、1つの文章内に限定されており、その文章内からリンクが張られた別の文章までを1つの文章として扱ってサーチするものではない。大規模な文章を階層化して複数の文章に分割することで、該文章が一画面に表示されたときに一覧性に優れた文章として、たとえばインターネットで公開される文章があるが、このような文章から目的の文章を効率よく探すには、階層化して分 50

割されたそれらの文章を本来の形の1つの文章としてまとめ扱うことが重要である。

【0005】特開平8-249342号公報には、ハイパーテキストを構成するカードの中からリンク情報に従って検索表インデックスを参照し、指示した語句を含むカード(文章)およびそのカードをリンク先に持つカードを検索する装置が示される。しかしながら、本公報に開示の技術ではリンクされている文章(カード)を1つに結合してから検索するものではないから、検索のヒット率に優れない。

【0006】それゆえに、この発明の目的は検索のヒット率に優れる文章検索装置および方法ならびに文章検索プログラムを記録したコンピュータで読取り可能な記録媒体を提供することである。

【0007】

【課題を解決するための手段】請求項1に記載の文章検索装置は、各データ内で相互に関連付けするためのリンク情報に従い関連付けられた文章データを含む複数のデータを記憶する記憶手段と、記憶手段中の複数のデータをリンク情報に従い統合処理して1つ以上の統合文章データに生成するための統合手段と、1つ以上の検索文字列データを指定する指定手段と、統合手段により統合された1つ以上の統合文章データのそれぞれにおいて、指定手段により指定された1つ以上の検索文字列データが含まれるか否かを検索する検索手段と、検索手段の検索結果を表示する表示手段とを備える。

【0008】したがって、小さな画面での見易さを優先してリンク情報で階層化されて相互に関連付けられている複数のデータを検索対象とする場合でも、これらの複数データは統合手段にて統合された文章データとされ、検索手段はこの統合文章データを対象にして指定された検索文字列データを検索する。

【0009】それゆえに、統合文章データを検索対象とするから、検索手段によるフルテキストサーチであっても不要なデータは検索されず、1つ以上の検索文字列データのAND/ORの検索で必要なデータは表示して不要なデータは表示しない事ができて、検索時のヒット率を高めることができる。

【0010】請求項2に記載の文章検索装置は、請求項1に記載の装置がリンク情報を記憶するリンク情報記憶手段をさらに有する。そして、統合手段は、リンク情報記憶手段中のリンク情報に基づいて複数データにおいて関連付けがループしているか検出するループ検出手段をさらに有して、ループ検出手段により複数データにおいてループしていると検出されたデータは統合処理の対象としない。

【0011】したがって、リンク情報による関連付けがループしていることはループ検出手段にて検出されて、統合処理の対象とされないから統合処理が該ループにより無限ループ処理となることが回避されて、検索の無駄

がなくなる。

【0012】請求項3に記載の文章検索装置は、請求項2に記載の文章検索装置の統合手段は、ループ検出手段によりループしていると検出された場合には、ループしている関連付けにおける1つ以上の部分で関連付けが無しとされた後のリンク情報に基づいて、複数データについて統合処理をする。

【0013】したがって、ループしている1つ以上の部分で関連付けを断つ(切る)ことで、ループを解消して断った部分について個々に統合処理して複数の統合文章データを作成できる。また、統合処理が該ループにより無限ループ処理となることが回避されて、検索の無駄がなくなる。

【0014】請求項4に記載の文章検索装置は、請求項2または3に記載の文章検索装置において統合手段は、リンク情報記憶手段中のリンク情報に基づいて複数データにおいて一方の文章データから他方の文章データへの関連付けを判定する関連付け判定手段をさらに有する。そして、関連付け判定手段により一方文章データ中の複数の単語のそれぞれから他方文章データ中の複数の単語のそれぞれに関連付けされていると判定された場合には、統合手段は他方文章データを前記関連付けに基づいて分割して、分割後の文章データのそれぞれを用いて統合処理する。

【0015】したがって、リンク情報に従う関連先の文章データの関連部分のみを統合文章となる該リンク情報に従う関連元の文章データに取込むことができるから、統合文章データは必要とされるデータのみを用いて生成されて、検索手段による検索に無駄がない。

【0016】請求項5に記載の文章検索装置は、請求項4に記載の文章検索装置において関連付け判定手段により、一方文章データ中の複数の単語から他方文章データ中の同じ単語に関連付け去れていると判定された場合には、統合手段は他方文章データをそのまま一方文章データ中に取込んで統合処理をする。

【0017】したがって、リンク情報に従う関連先の文章データそのままを統合文章となる該リンク情報に従う関連元の文章データに取込むことができるから、統合文章データは検索に必要とされるデータを用いて生成されて、検索手段による検索の成功率が高まる。

【0018】請求項6に記載の文章検索装置は、請求項1ないし5のいずれかに記載の文章検索装置の統合処理において、文章データが統合される場合には、関連付けにおける関連元の文章データ中に対応の関連先の文章データが段落を下げて取込まれる。

【0019】したがって、表示手段による検索結果で統合文章データが表示されるとき、または検索手段の検索時に、段落が下げられていることによりリンク情報に従う関連先の文章データであることが容易に検出できる。

【0020】請求項7に記載の文章検索装置は、請求項

1ないし6のいずれかに記載の文章検索装置の統合処理において、文章データが統合される場合に、関連付けにおける関連元の文章データ中の対応する関連先文章データの取込位置の文字列と該関連先文章データの先頭文字列が同じであるときは、一方の文字列が省略される。

【0021】したがって、統合処理では、統合文章データにおいて同一文字列が連続して位置することがないような編集が併せて行われるから、統合文章データは内容が変更されることなく冗長な部分は排除されて、検索手段による検索効率が高まる。

【0022】請求項8に記載の文章検索装置は、請求項1ないし7のいずれかに記載の文章検索装置の表示手段において、検索結果である統合文章データはアウトライン機能モードで表示される。

【0023】したがって、表示される統合文章データをアウトライン機能によりインデント処理などしてユーザに対して見やすい態様で表示できる。

【0024】請求項9に記載の文章検索装置は、請求項1ないし8のいずれかに記載の文章検索装置の統合処理において、関連先のデータが文章データでない場合に、該データは処理対象とされない。

【0025】したがって、文章データでないデータ、たとえば画像データまたは音声データは統合されないようにすることができて、統合文章データに検索可能でないデータが取り込まれて検索処理が誤動作することは回避される。また、表示手段による検索結果としての統合文章データ表示の処理が誤動作することが回避される。

【0026】請求項10に記載の文章検索装置は、請求項1ないし8のいずれかに記載の文章検索装置の統合処理において、関連先のデータが文章データでない場合には、該データを識別するために予め付与された文字列データを用いて処理される。

【0027】したがって、文章データでないデータ、たとえば画像データまたは音声データは、それを識別するためのファイル名またはデータ名の文字列データで統合するようにできて、これらファイル名またはデータ名を検索文字列として指定できる。

【0028】請求項11に記載の文章検索装置は、請求項1ないし10のいずれかに記載の文章検索装置の記憶手段中の複数データは、ハイパーテキストを構成するデータである。

【0029】したがって、統合文章データとしてハイパーテキストによるものを得ることができて、これを対象にして検索できる。

【0030】請求項12に記載の文章検索装置は、各データ内で相互に関連付けるためのリンク情報に従い関連付けられた文章データを含む複数のデータを記憶する記憶手段と、1つ以上の検索文字列データを指定する指定手段と、記憶手段中の複数データ中から指定手段により指定された検索文字列データを検索する検索手段と、検

索手段による検索結果を表示する表示手段とを備える。そして、検索手段は、データの検索中にリンク情報があれば、該データの検索を中断し該リンク情報に従う関連先のデータを検索して検索終了すれば、該リンク情報に従う関連元のデータの検索を再開しながら、1つ以上の検索文字列データが検索された文章データを逐次統合して統合文章データに生成する。

【0031】したがって、データ検索時にリンク情報があれば、該リンク情報に従う関連先のデータもその時点で検索することで、指定された1つ以上の検索文字列データのAND/ORなどの検索に必要なデータは逐次統合文章データとして取込み、不要なデータは取込みされないようできる。それゆえに、検索文字列を用いた検索と統合文章データの作成とを並行して行なうことができる。

【0032】請求項13に記載の文章検索方法は、各データ内で相互に関連付けるためのリンク情報に従い関連付けられた文章データを含む複数のデータを対象にした文章検索方法であって、複数のデータを前記リンク情報に従い統合処理して1つ以上の統合文章データに生成するための統合ステップと、1つ以上の検索文字列データを指定する指定ステップと、統合ステップにより統合された1つ以上の統合文章データのそれぞれにおいて、指定ステップにより指定された1つ以上の検索文字列データが含まれるか否か検索する検索ステップと、検索ステップの検索結果を表示する表示ステップとを備える。

【0033】したがって、小さな画面での見易さを優先してリンク情報で階層化されて相互に関連付けられている複数のデータを検索対象とする場合でも、これらの複数データは統合手段にて統合された文章データとされ、検索手段はこの統合文章データを対象にして指定された検索文字列データを検索する。

【0034】それゆえに、統合文章データを検索対象とするから、検索手段によるフルテキストサーチであっても不要なデータは検索されず、1つ以上の検索文字列データのAND/ORの検索に必要なデータは表示して不要なデータは表示しない事ができて、検索時のヒット率を高めることができる。

【0035】請求項14に記載の文章検索方法は、各データ内で相互に関連付けるためのリンク情報に従い関連付けられた文章データを含む複数のデータを対象にした文章検索方法であって、1つ以上の検索文字列データを指定する指定ステップと、複数データ中から指定ステップにより指定された検索文字列データを検索する検索ステップと、検索ステップによる検索結果を表示する表示ステップとを備える。そして、検索ステップは、データの検索中にリンク情報があれば、該データの検索を中断し該リンク情報に従う関連先のデータを検索して検索終了すれば、該リンク情報に従う関連元のデータの検索を再開しながら、1つ以上の検索文字列データが検索され

た文章データを逐次統合して統合文章データに生成する。

【0036】したがって、データ検索時にリンク情報があれば、該リンク情報に従う関連先のデータもその時点で検索することで、指定された1つ以上の検索文字列データのAND/ORなどの検索に必要なデータは順次統合文章データとして取込み、不要なデータは取込みされないようできる。それゆえに、検索と統合文章データ作成とを並行して行なうことができる。

【0037】請求項15に記載の記録媒体は、各データ内で相互に関連付けるためのリンク情報に従い関連付けられた文章データを含む複数のデータを対象にした文章検索方法をコンピュータに実行させるための文章検索プログラムを記録したコンピュータで読取り可能な記録媒体である。

【0038】そして、この文章検索方法は、複数のデータをリンク情報に従い統合処理して1つ以上の統合文章データに生成するための統合ステップと、1つ以上の検索文字列データを指定する指定ステップと、統合ステップにより統合された1つ以上の統合文章データのそれぞれにおいて、指定ステップにより指定された1つ以上の検索文字列データが含まれるか否かを検索する検索ステップと、検索ステップの検索結果を表示する表示ステップとを備える。

【0039】したがって、小さな画面での見易さを優先してリンク情報で階層化されて相互に関連付けられている複数のデータを検索対象とする場合でも、これらの複数データは統合手段にて統合された文章データとされ、検索手段はこの統合文章データを対象にして指定された検索文字列データを検索する。

【0040】それゆえに、統合文章データを検索対象とするから、検索手段によるフルテキストサーチであっても不要なデータは検索されず、1つ以上の検索文字列データのAND/ORの検索に必要なデータは表示して不要なデータは表示しない事ができて、検索時のヒット率を高めることができる。

【0041】請求項16に記載の記録媒体は、各データ内で相互に関連付けるためのリンク情報に従い関連付けられた文章データを含む複数のデータを対象にした文章検索方法をコンピュータに実行させるための文章検索プログラムを記録したコンピュータで読取り可能な記録媒体である。そして、この文章検索方法は、各データ内で相互に関連付けるためのリンク情報に従い関連付けられた文章データを含む複数のデータを対象にした文章検索方法であって、1つ以上の検索文字列データを指定する指定ステップと、複数データ中から指定ステップにより指定された検索文字列データを検索する検索ステップと、検索ステップによる検索結果を表示する表示ステップとを備える。

【0042】そして検索ステップは、データの検索中に

リンク情報があれば、該データの検索を中断し該リンク情報に従う関連先のデータを検索して検索終了すれば、該リンク情報に従う関連元のデータの検索を再開しながら、1つ以上の検索文字列データが検索された文章データを逐次統合して統合文章データに生成することを特徴とする。

【0043】したがって、データ検索時にリンク情報があれば、該リンク情報に従う関連先のデータもその時点で検索することで、指定された1つ以上の検索文字列データのAND/ORなどの検索に必要なデータは順次統合文章データとして取込み、不要なデータは取込みされないようできる。それゆえに、検索と統合文章データの作成とを並行して行なうことができる。

【0044】

【発明の実施の形態】図1はこの発明の実施の形態に適用される文章検索装置のブロック構成図である。図1において文章検索装置は該装置全体を制御するCPU（中央処理装置の略）11、検索する文字列を入力したり検索結果を選択するために操作されるキーボードなどの入力部12、インターネットなどの通信ネットワークからデータを入手するための通信部13、通信ネットワークから収集し統合されたデータを保存するためのHDD（ハードディスクドライブ）などの記憶部14、記憶された内容の中から指定された文字列を検索するための検索部15、入力部12から入力された文字列や通信部13で入手したデータを表示するためのCRT（陰極線管の略）や液晶ディスプレイなどの表示部16、検索結果得られたデータを記憶する候補バッファ17、選択された文章をどのモードで表示するかを記憶する表示モード記憶領域18、表示／編集対象である文章を格納するためのテキストバッファ19、収集した文章を格納するテキスト記憶領域20、後述するリンクの相関関係を記憶するための相関記憶部27、収集した文章を統合した統合文章に加工した後保存するための統合文章記憶領域21、検索途中のアドレスをセーブするアドレスセーブバッファ22、検索中の文章に対応してフラグを割当てる検索フラグバッファ23、読出されて検索された文章にフラグを設定する読出フラグバッファ24、リンク先の文章を取込むときの段落の深さを記憶するための取込深さ記憶領域25からなる。

【0045】また、後述する各種処理フローチャートをプログラムとして格納したFD（フレキシブルディスク）などの記録媒体26が該文章検索装置のために準備され、記録媒体26から図示されないFDD（フレキシブルディスクドライバ）などを介してプログラムが読出されて該装置のメモリに格納されCPU11の制御の下に実行される。

【0046】本文章検索装置では、個人のHD（ハードディスク）内の情報やLAN（ローカルエリアネットワーク）、インターネット上からテキスト情報が収集さ

れ、該装置内に蓄積され、蓄積されているハイパーテキストのデータがリンク関係に基づき1つ以上の文章に統合される。そして、ユーザからの検索指示が検索文字列入力により行なわれると、1つ以上の統合文章から入力された検索文字列が検索されて、該検索文字列を含む統合文章が指定された表示形態で表示される。

【0047】図2は、この発明の実施の形態に適用されるハイパーテキストの第1例を示す図である。図2では文章Aから文章Bと文章Cに同時にリンクが張られている。データの個数が多い場合には、データを階層化して格納し階層化された関係であるリンク元とリンク先の関係(親子関係)をリンクで結ぶことで、一覧性をよくしている。

【0048】図3(A)～(C)は図2の各文章の内部データを示す図である。この例では、インターネットで一般的に使用されているものが示されており、HTML(Hyper Text Mark-up Languageの略)のバージョンにより、サポートされているタグも変わってくる。リンク情報は場所(検索アドレスまたは記憶アドレス)とファイル名で指定される。図では“<HTML>”から“</HTML>”までがブラウザで見るデータで、“<BODY>”から“</BODY>”が本文で、ブラウザで画面内に表示されるものである。“<BR>”が改行で通常のエディタの改行と同じように文字列が折り返して表示される。“< HREF = “\*\*\*. HTML #\*\*\*”>”から“</A>”までが、リンクを持っていて“\*\*\*. HTML”というファイルの“\*\*\*”という名前の位置を指していることを示している。“#\*\*\*”がなければ“\*\*\*. HTML”というファイル全体を指していることになる。“\*\*\*”という名前はデータ内に“<A NAME = “\*\*\*”></A>”で示される。“<TITLE>”から“</TITLE>”までがタイトルで通常ブラウザのウィンドウの一番上に表示される。

【0049】図4は図3(A)～(C)のデータが統合された結果の統合文章を示す図である。図の統合文章ATではリンク先の文章はリンク元の文章において段落を下げて取込まれるので見やすくなっている。この統合文章ならば今までのサーチエンジンで検索可能となる。元の文章A～Cから文字列“奈良県”と文字列“大和郡山市”のANDの条件で検索すると、どの文章も条件に該当せず一致しないが、図4の統合文章ATでは文字列“奈良県”と文字列“大和郡山市”の両方の文字列が同時に含まれるので検索することが可能となる。

【0050】図5は、この発明の実施の形態に適用されるハイパーテキストの第2例を示す図である。文章Dから文章Eへ、文章Eから文章Fへ、そして文章Fから文章Dへそれぞれリンクが張られて、ループしている例が示される。図6(A)～(C)は、図5の各文章の内部データを示す図である。ここではリンク情報は場所とフ

ファイル名で指定される。

【0051】検索時には、図5に示されるようなリンクによるループのチェックをしなければ検索処理が無限ループ処理に陥ってしまう。そこで、既に読出して検索した文章には読出フラグをセットして、検索しようとする文章に対応の読出フラグが既にセットされていればリンクがループしていると判断され、そこでそのリンクを検索するのを中止する。またループしていればループを切る場所を順番に変えることで、複数種類の統合文章を作成することができる。この、文章に対応して読出フラグを格納する領域は図1の読出フラグバッファ24である。

【0052】図7～図9は、図5および図6に示される各文章が統合された結果の統合文章の第1～第3例を示す図である。図10～図12は、図5および図6に示される各文章が改良して統合された統合文章の第1～第3例を示す図である。

【0053】図7～図9では統合された結果文字列が重複して見にくくなるので、リンク先の文章の先頭の文字列がリンク元の文章の文字列と一致していれば、リンク先の先頭の文字列をカット(省略)して統合することで図10～図12に示されるように見やすい統合文章に改良することができる。図10～図12の統合文章ならば今までのサーチエンジンで検索可能となる。またどれを元にするか見やすいものを選ぶことができる。

【0054】図13は、この発明の実施の形態に適用されるハイパーテキストの第3例を示す図であり、同一文章内の複数の異なる位置にリンクが張られている例が示される。文章G内の2ヵ所からそれぞれ文章Hの1番目に位置する文字列“ブルドック”と文章Hの2番目に位置する文字列“シャム猫”にリンクが張られている。図14(A)と(B)は、図13の各文章の内部データを示す図である。リンク情報は場所とファイル名で指定される。図15は図13に示される文章が統合された結果の統合文章を示す図である。リンク先の文章の必要な部分(文字列)だけがリンク元の文章のリンク位置に統合されているので見やすくなっている。

【0055】図16は、この発明の実施の形態に適用されるハイパーテキストの第4例を示す図であり、同一文章の同じ位置の文字列(名前)に、複数の異なるリンクが張られている例である。図16では文章K内の2ヵ所から同一文章L内の1番目に位置する文字列“家畜”にリンクが張られている。図17(A)と(B)は図16の各文章の内部データを示す図である。リンク情報は場所とファイル名で指定される。

【0056】図18は、図17(A)と(B)のデータが統合された結果の統合文章を示す図である。この統合文章KTならば今までのサーチエンジンで検索可能となる。データ(文字列)が重複して格納されているが、検索抽出したときに表示されるとリンク先の内容がリンク

元の文章に入っているので見やすくなる。

【0057】図19はこの発明の実施の形態に適用されるデータ収集時の処理フローチャートである。図20は、図19の処理において表示されるデータ取込条件設定画面例を示す図である。

【0058】まず、データ収集時には図20の収集対象となるデータの取込条件のための指定ウィンドウが表示部16に表示されて(S101)、ユーザが入力部12を操作して日本語/英語などの収集対象となる文章の言語、収集の頻度であるチェック頻度などの収集方法を指定すると(S102)、検索先であるインターネットなどのアドレスが初期化されて(S103)、アドレスに基づいて通信部13を介して順次、情報検索され(S104)、画像/音声などのデータを省いた文章データのみが記憶部14に読込まれる(S105)。

【0059】ここで、インターネットのアドレスは現在4バイトで表現されている。0.0.0.0から256.256.256.256まで表現可能であり、一部のアドレスを除き重複せずに割振られている。このような値を順次カウントアップしながらアドレス更新して更新されたアドレスに基づいて情報検索して文章データを取込むことができる。

【0060】図21は、この発明の実施の形態によるリンク先がテキスト以外の例を示す図である。図21のようにリンク先が画像や音声などのデータの場合はそのファイル名等の文字列データも取込む。取込まれたデータをHDなどの記憶部14に保存し(S106)、検索するアドレスをアップし(S107)、図20で予め入力されて定められた条件に一致するアドレスをすべて検索終了したかチェックし(S108)、検索終了でなければ再度検索(S104)から繰返される。

【0061】なお、記憶部14に保存されたデータは、後述される検索処理のためにテキスト記憶領域20に格納される。

【0062】図22は、この発明の実施の形態に適用されるデータ統合時の処理フローチャートである。図23は図22のテキスト統合の詳細処理フローチャートである。まず、インターネットを検索してテキスト記憶領域20に格納している文章データ検索のアドレスを初期化し(S111)、図23のテキスト統合ルーチンをコールし(S112)、その後アドレスをアップし(S113)、終了かチェックし(S114)、終了ならば統合文章を統合文章記憶領域21に保存する(S115)。終了でなければS112の処理から続ける。

【0063】図23のテキスト統合ルーチンではS111またはS113で指定されたアドレスが指す文章をテキスト記憶領域20から読出し(S116)、読出フラグバッファ24において読出された文章に対応の読出フラグがセットされているかチェックし(S117)、読出フラグがセットされていれば既に読込まれているの

で、この文章の処理を終える。一方、読出フラグがセットされていなければ読出フラグをセットし(S118)、その文章に含まれているリンクを順番に読出し(S119)、各リンク先の文章に対応の読出フラグがセットされているかチェックし(S120)、読出フラグがセットされていれば、既に該リンク先の文章は記憶領域21に読込まれているので、このリンク先の文章の処理を終える。

【0064】読出フラグがセットされていなければ読出フラグをセットし(S121)、記憶領域21の現在のポイント位置より段落を下げて記憶領域21に取込んで(S122)、リンク先の文章をリンク元の文章に統合する(S123)。ここで、リンク先の文章の先頭に位置する文字列とリンク元の文章の取込み位置の文字列が一致していれば、前述したように一致している文字列の一方をカットして統合することも可能である。

【0065】図24(A)と(B)は、リンク先のデータを取込んで統合するか否かによる統合文章の例を示す図である。リンク先が画像などの文章以外のデータならば図24(B)に示されるようにそのファイル名またはタイトル名を取込むことも可能であり、図24(A)のように取込まないこともできる。この取込要/否の設定は予め入力部12を介してユーザが設定すればよい。

【0066】このテキスト統合ルーチンはこのフローで示されている処理全体を行なう(リカーシブルコール)ため、その統合先の文章におけるリンクも処理してリンク先の文章を記憶領域21に取込むことになる。リンク先文章の処理を終了するとリンク元文章に戻って記憶領域21における取込位置を示す段落を元の位置に戻す(S124)。

【0067】続いて、次のリンク先を検索し(S125)、まだリンクがあればその文章を検索して読出すことから処理を続ける。もうリンクがなければ(S126)、リンク元文章の取込位置の文字列とリンク先文章の取込まれる先頭文字列が同じかチェックし(S127)、同じならば同じ文字列の一方をカットし(S128)統合文章に追加し(S129)、リターンする。

【0068】図25(A)～(D)は、図2、図5、図13および図16のハイパーテキストにおける相関関係を表形式で示す相関表の図である。この相関表は図1の相関記憶部27に記憶される。

【0069】まず、ハイパーテキストにおけるリンク関係が複雑な場合には、予めリンク関係を表形式でまとめて、この表を参照しながら文章を統合すると、ループしているリンクを検索対象から外したり、ループの切る位置を順次変更して異なる統合文章を作成したり、同一文章の複数の単語にリンクされているときに文章を分割したりすることが容易に可能となる。

【0070】図25(A)～(D)の相関表ではリンク元文章のファイル名と、そのリンク元の単語(文字列)



と、リンク先文章のファイル名と、そのリンク位置が格納される。このリンク位置はリンク先単語（文字列）または該単語の該文章における位置を示す。

【0071】図26はこの発明の実施の形態による相関表を作成するための処理フローチャートである。まず相関記憶部27の相関表が初期化され（S211）、検索のためのアドレスが初期化され（S212）、検索アドレスが指す記憶部14の文章のファイルが読出される（S212）。

【0072】そして、読出されたファイルである文章に関する検索のためのテキストポインタが初期化され（S214）、このテキストポインタを用いて読出された文章内が検索され（S215）、検索結果得られた情報はリンクかチェックされ（S216）、リンクならば図25（A）～（D）に示されるように相関表にリンク元ファイル名、リンク元単語、リンク先ファイル名およびリンク先位置がセットされ（S217）、該文章のテキストポインタを用いた検索が終了するまで繰返される（S218）。

【0073】該文章のテキストポインタを用いた処理が終了すれば前述の検索のアドレスが更新され（S219）、更新された検索アドレスを用いてテキスト記憶領域20中で別の文章であるファイルがあるか検索されてチェックされる（S220）。なければ相関表において同じファイル名がリンク元ファイル名およびリンク先ファイル名として出現してループしているかチェックされる（S221）。

【0074】図25（B）の場合、文書Dがリンク元ファイル名およびリンク先ファイル名として出現してループしているからループ位置を順次変えて取込み、別々の統合文章を作成したり（S222）、ループしている部分を取込まないようにすることもできる。

【0075】また相関表に同じファイルの組合せがあるかチェックし（S223）、あればリンク元もリンク先も異なるファイルかチェックし（S224）、異なれば位置の後のリンク先の前後で分割して取込み統合文章を作成する（S225）。これは、図25（C）の場合であり、図15の統合文章が得られる。

【0076】またリンク元が異なるがリンク先が同じかチェックし（S226）、同じならばリンク先の文章をすべて取込む（S227）。これは、図25（D）の場合であり図18の統合文章が得られる。

【0077】図27はこの発明の実施の形態による統合文章検索時のフローチャートである。図28は図27の処理にて表示される検索文字列入力ウィンドウの一例を示す図である。図29は図27の処理にて表示される表示モード指定画面の一例を示す図である。

【0078】図27においてまず、図28の検索文字列入力ウィンドウが表示部16に表示されて（S131）、ユーザが1つ以上の検索文字列を入力部12から

入力すると（S132）、1つ以上入力された文字列はたとえばAND条件で検索される。そしてまず統合文章記憶領域21の先頭に登録された統合文章にポインタが移動し（S133）、候補バッファ17を初期化し（S134）、ポインタが示す統合文章記憶領域21中の統合文章を、予め入力された検索文字列を用いて検索部15によりフルテキストサーチエンジンなどで検索する（S135）。

【0079】このときに、文章以外のデータ（画像データまたは音声データ）も取込むよう予め指定されているならば、図24（B）のように画像データまたは音声データなどのファイル名を取込むために、リンク先のファイル名も検索される。

【0080】入力されて指定されていた検索文字列がすべて該統合文章内に存在していれば（AND条件成立ならば）、候補バッファ17にその統合文章名（リンクのルートとなる文章名）をセットし（S137）、候補バッファ17のポインタが更新される（S138）。そして、記憶領域21のポインタを次に登録されている統合文章に移動し（S139）、統合文章がまだ存在するかどうかチェックされ（S140）、まだ統合文章があれば、検索（S135）の処理から繰返される。

【0081】統合文章がなければ、検索文字列を用いた検索の結果、得られた統合文章名が記憶されている候補バッファ17の内容が表示部16に表示される（S141）。ユーザは、表示部16の表示内容を見て呼出した統合文章名を入力部12を介して指定し（S142）、図29の画面にて表示モードも指定すると（S143）、指定された表示モードで指定された統合文章が表示部16に表示される（S144）。

【0082】図30は、この発明の実施の形態による複数の単語を検索するときの例を説明する図である。リンク元になる文章を検索する時点で、図30に示される文章のファイル名をすべて初期化し、対応する検索単語が見つかったらそのファイル名を書込み、リンクをすべて検索した時点で、検索単語のそれぞれについてファイル名が書込まれていれば、複数の単語（検索文字列）を用いたAND検索ができたことになり、1つでも書込まれていればOR検索ができたことになる。

【0083】図31（A）と（B）は、この発明の実施の形態による表示モードによる表示の違いを示している図である。テキストモード指定のときは文字列が単純に羅列されている。アウトラインモード指定のときはアウトラインの表示モードにより段落になったり、リンク先が表示されない。図の場合は、段落深さ2のレベルの文字列が表示されていない。

【0084】図32はこの発明の実施の形態による統合文章のアウトライン表示のフローチャートである。図32において、まず表示モード記憶領域18から予めユーザにより設定された表示モードを取出し（S191）、

ユーザが選んだ統合文章を読み込み(S192)、統合文章のポインタを初期化し(S193)、表示用のポインタを初期化し(S194)、表示している段落深さのデータを初期化する(S194)。そして統合文章のポインタが指している1行を取出し(S196)、統合文章のポインタを更新し(S197)、設定表示モードが通常(テキスト)モードかチェックし(S198)、通常モードならばその1行を表示し(S201)、表示ポインタを更新する(S202)。

【0085】設定表示モードがアウトラインモードならば、文章のインデント位置を取出し(S199)、そのインデント位置が図29で予め指定される表示の段落深さと比較して(S200)、小さければ上のレベルなので、その1行を表示し(S201)、表示ポインタを更新する(S202)。この操作を該統合文章が終了するまで繰返す(S203)。

【0086】上述した文章の検索処理では、複数の文章を検索して取込んだ後1つの文章に統合する処理を行ってから、統合文章について従来のフルテキストサーチエンジンを用いて検索を行っていたが、文章検索時にリンクが出てきたときは、その時点で検索をリンク先の文章に切替えるようにしてもよい。

【0087】図33は、この発明の実施の形態による検索時リンク追跡のフローチャートである。まず、図28のような検索ウィンドウが表示され(S151)、ユーザが1つ以上の検索文字列を入力すると(S152)、テキスト記憶領域20の先頭の文章にポインタが移動し(S153)、候補バッファ17を初期化し(S154)、検索フラグバッファ23で該文章に対応の検索フラグがセットされているかチェックし(S155)、セ

ットされていればこの文章は既に検索されているから検索を終了する。

【0088】そしてポインタが指す次の文章を検索する(S156)。このときに、文章以外も取込むような指定がされているならば図24(B)のようにファイル名を取込むことができるようリンク先のファイル名も検索する。リンクがあれば(S157)現在の文章におけるアドレスをアドレスセーブバッファ22にセーブし(S158)、この文章に対応の検索フラグバッファ23における検索フラグをセットし(S259)、リンク先の

文章を検索する(S160)。

【0089】リンク先が検索終了すれば検索フラグバッファ23における対応の検索フラグをリセット(クリア)し(S171)、セーブしていたアドレスセーブバッファ22中のアドレスから検索を再開する(S172)。

【0090】予め入力された検索文字列がすべてあれば(S173)、候補バッファ17にリンク元の文章名を統合文章名としてセットし(S174)、候補バッファ17のポインタを更新し(S175)、検索のためのテ

キスト記憶領域20のポインタを次の文章に移動し(S176)、文章がまだ存在するか否かチェックし(S177)、あれば検索フラグのチェックから繰返し、なければ候補バッファ17の内容である1つ以上の統合文章名を表示し(S178)、ユーザが呼出したい統合文章名が指定される(S179)。この指定に応じてCPU11は指定統合文章名であるリンク元文章と、これにリンクに従い関連付けられている他の文章を統合して統合文章をテキストバッファ19に作成する。表示モード(図29参照)も指定されると(S180)、指定された表示モードで作成されたテキストバッファ19の統合文章が表示される(S181)。

【0091】この処理フローチャートでは、検索エンジンにリンク先の検索も行なう対応を入れなければならないが、統合された文章データを内部にて蓄えなくて良いという利点がある。

【0092】今回開示された実施の形態はすべての点で例示であって制限的なものではないと考えられるべきである。本発明の範囲は上記した説明ではなくて特許請求の範囲によって示され、特許請求の範囲と均等の意味および範囲内でのすべての変更が含まれることが意図される。

【図面の簡単な説明】

【図1】この発明の実施の形態に適用される文章検索装置のブロック構成図である。

【図2】この発明の実施の形態に適用されるハイパーテキストの第1例を示す図である。

【図3】(A)～(C)は図2の各文章の内部データを示す図である。

【図4】図3(A)～(C)のデータが統合された結果の統合文章を示す図である。

【図5】この発明の実施の形態に適用されるハイパーテキストの第2例を示す図である。

【図6】(A)～(C)は、図5の各文章の内部データを示す図である。

【図7】図5および図6に示される各文章が統合された結果の統合文章の第1例を示す図である。

【図8】図5および図6に示される各文章が統合された結果の統合文章の第2例を示す図である。

【図9】図5および図6に示される各文章が統合された結果の統合文章の第3例を示す図である。

【図10】図5および図6に示される各文章が改良して統合された統合文章の第1例を示す図である。

【図11】図5および図6に示される各文章が改良して統合された統合文章の第2例を示す図である。

【図12】図5および図6に示される各文章が改良して統合された統合文章の第3例を示す図である。

【図13】この発明の実施の形態に適用されるハイパーテキストの第3例を示す図である。

【図14】(A)と(B)は図13の各文章の内部デー

タを示す図である。

【図15】図13に示される文章が統合された結果の統合文章を示す図である。

【図16】この発明の実施の形態に適用されるハイパーテキストの第4例を示す図である。

【図17】(A)と(B)は図16の各文章の内部データを示す図である。

【図18】図17(A)と(B)のデータが統合された結果の統合文章を示す図である。

【図19】この発明の実施の形態に適用されるデータ収集時の処理フローチャートである。

【図20】図19の処理において表示されるデータ取込条件設定画面例を示す図である。

【図21】この発明の実施の形態によるリンク先がテキスト以外の例を示す図である。

【図22】この発明の実施の形態に適用されるデータ統合時の処理フローチャートである。

【図23】図22のテキスト統合の詳細処理フローチャートである。

【図24】(A)と(B)は、リンク先のデータを取込んで統合するか否かによる統合文章の例を示す図である。

【図25】(A)～(D)は、図2、図5、図13および図16のハイパーテキストの相関関係を表形式で示す相関表の図である。

【図26】この発明の実施の形態による相関表を作成するための処理フローチャートである。

【図27】この発明の実施の形態に適用される統合文章検索時の処理フローチャートである。

【図28】図27の処理にて表示される検索文字列入力ウィンドウの一例を示す図である。

\*

\*【図29】図27の処理にて表示される表示モード指定画面の一例を示す図である。

【図30】この発明の実施の形態による複数の単語を検索するときの例を説明する図である。

【図31】(A)と(B)は、この発明の実施の形態による表示モードによる表示の違いを示している図である。

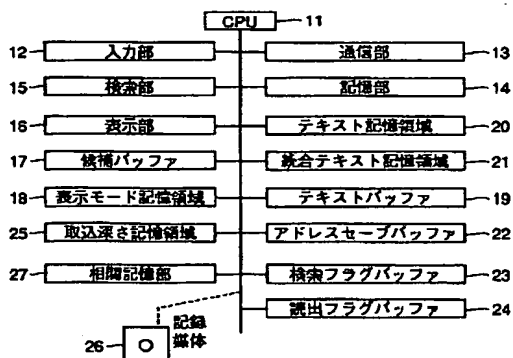
【図32】この発明の実施の形態による統合文章のアウトライン表示のフローチャートである。

【図33】この発明の実施の形態による検索時リンク追跡の処理フローチャートである。

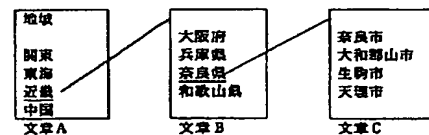
【符号の説明】

- 11 CPU
- 12 入力部
- 13 通信部
- 14 記憶部
- 15 検索部
- 16 表示部
- 17 候補バッファ
- 18 表示モード記憶領域
- 19 テキストバッファ
- 20 テキスト記憶領域
- 21 統合文章記憶領域
- 22 アドレスセーブバッファ
- 23 検索フラグバッファ
- 23 読出フラグバッファ
- 25 取込深さ記憶領域
- 26 記録媒体
- 27 相関記憶部

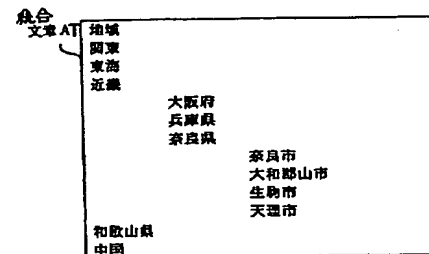
【図1】



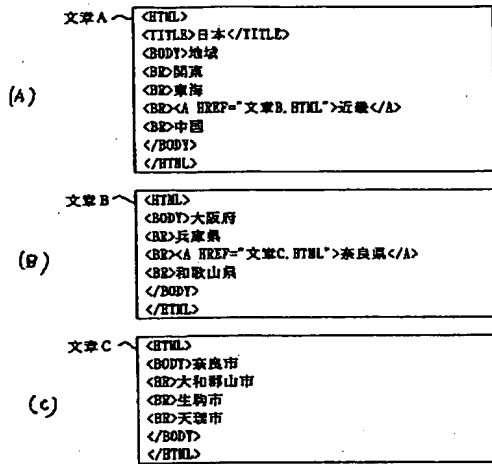
【図2】



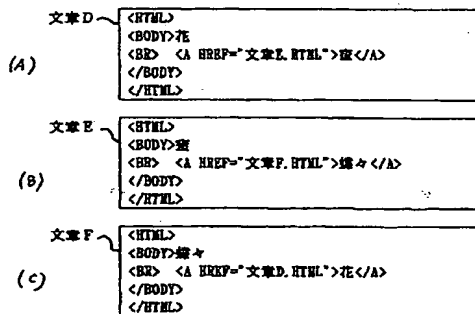
【図4】



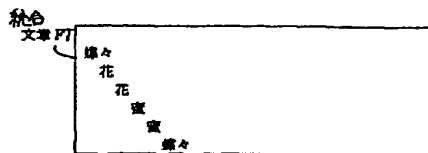
【図3】



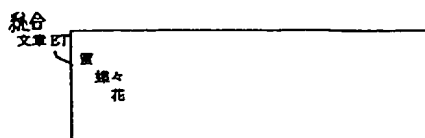
【図6】



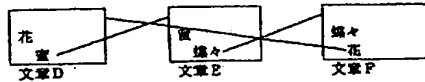
【図9】



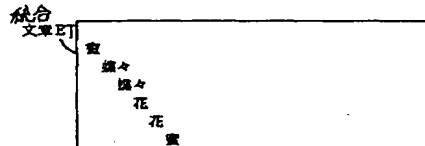
【図11】



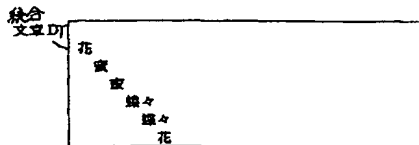
【図5】



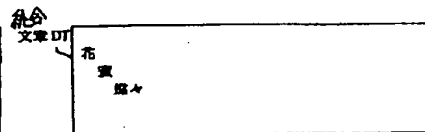
【図8】



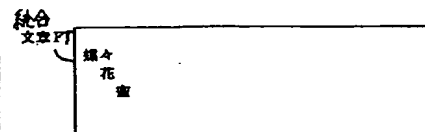
【図7】



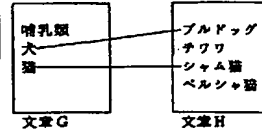
【図10】



【図12】



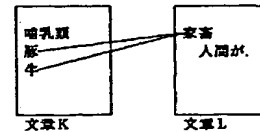
【図13】



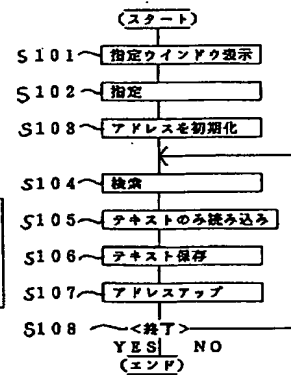
【図30】

検索単語	ファイル
A	文章A
B	文章A
C	文章B

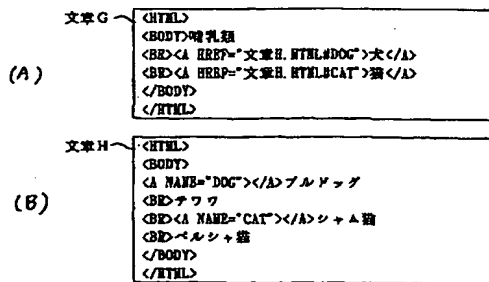
【図16】



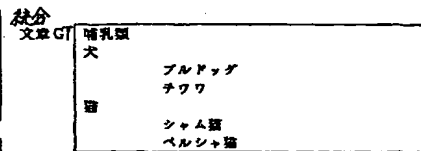
【図19】



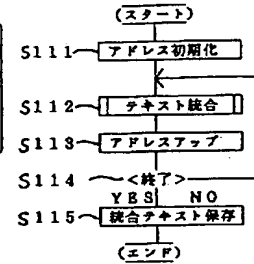
【図14】



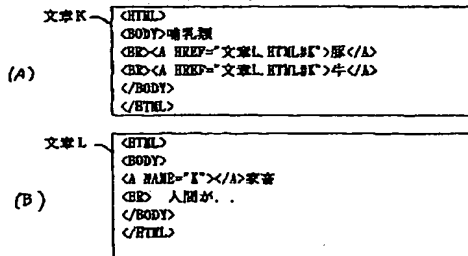
【図15】



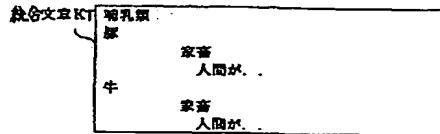
【図22】



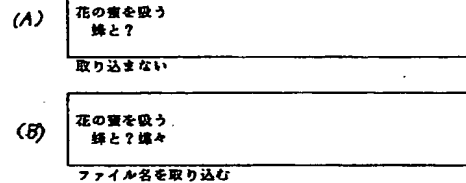
【図17】



【図18】



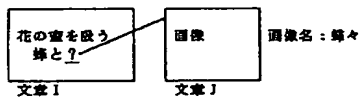
【図24】



【図20】

データ取り込み条件	
言語	●日本語 ○英語 ○ドイツ語
データチェック頻度	●1回/1日 ○1回/1週間 ○1回/1月 実行 中止

【図21】



【図28】

(A)

ファイル名	単語	リンク先	位置
文章A	近畿	文章B	先頭
文章B	奈良県	文章C	先頭
END			

(B)

ファイル名	単語	リンク先	位置
文章D	蜜	文章E	先頭
文章E	蜂々	文章F	先頭
文章F	花	文章D	先頭
END			

(C)

ファイル名	単語	リンク先	位置
文章G	犬	文章H	ブルドッグ
文章G	猫	文章H	シャム猫
END			

(D)

ファイル名	単語	リンク先	位置
文章K	豚	文章L	家畜
文章K	牛	文章L	家畜
END			

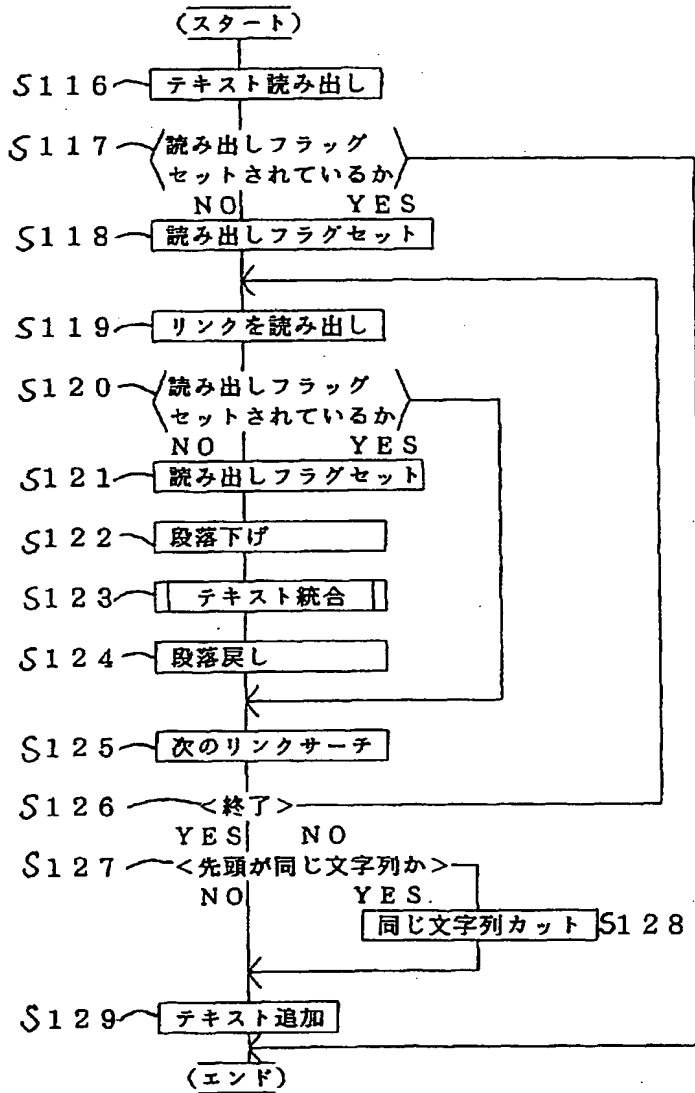
【図25】

検索文字	
文字列1:	
文字列2:	
文字列3:	
テキスト以外	
●取り込まない	○ファイル名を取り込む
実行	中止

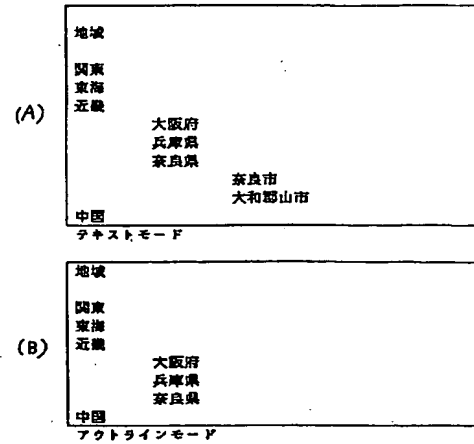
【図29】

表示指定	
○テキストモード	
●アウトラインモード	
深さ指定	2段
同一表示カット	
○する	●しない
実行	中止

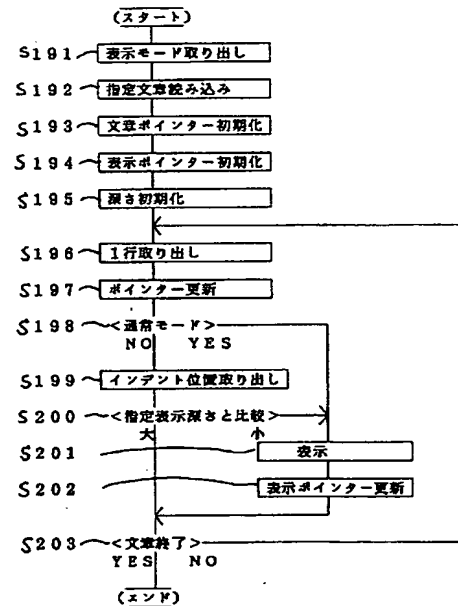
【図23】



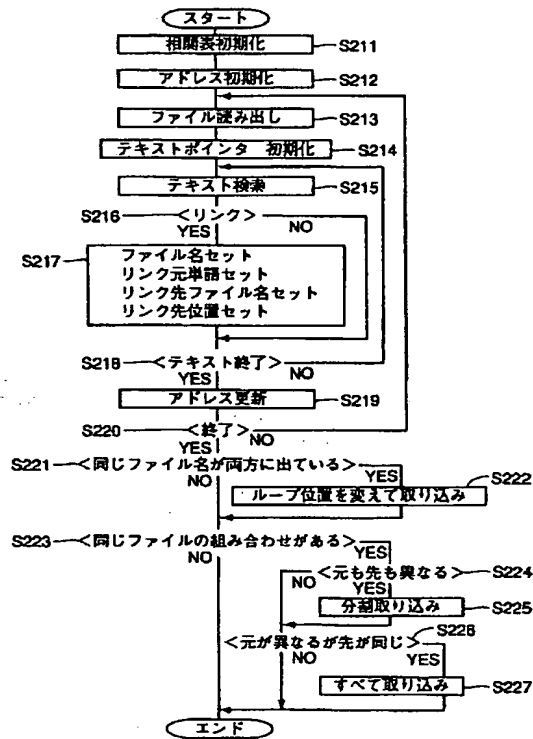
【図31】



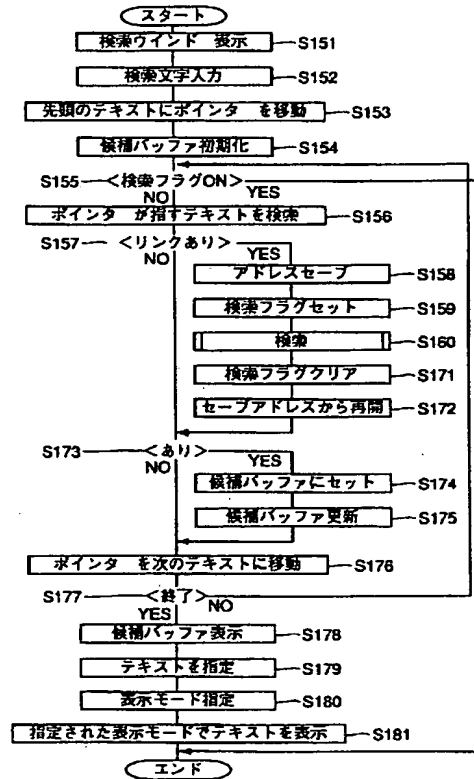
【図32】



【図26】



【図33】



【図27】

